A NOTE ON KAIMANN'S COEFFICIENT OF NETWORK

COMPLEXITY

by

Salah E. ELMAGHRABY †
Willy HERROELEN *

Onderzoeksrapport Nr. 7407.

† Department of Applied Economics, Katholieke Universiteit Leuven and EIASM, Brussels, on leave from North Carolina State University, Raleigh, N.C.

* Department of Applied Economics, Katholieke Universiteit Leuven.

ABSTRACT

In the area of Activity Networks several coefficients have been presented for measuring the "complexity" of a network. In this paper we discuss the Coefficient of Network Complexity forwarded recently by Kaimann. We argue that this coefficient, together with similar ones presented in the literature, lack some essential prerequisites to serve as universal measures. In addition two new measures of complexity are suggested : one that may be used when the objective of the analysis is the calculation of the critical path length ; the other to be used when determining the probability distribution function of terminal events.

A NOTE ON KAIMANN'S COEFFICIENT

OF NETWORK COMPLEXITY

In a recent paper Kaimann [3] proposes a new measure of the "complexity" of a network. This elicits from us the following comments, some of which are specific to his paper and some are on the general topic treated in it.

I. SPECIFIC COMMENTS.

(1) Unfortunately, the author never defines the term "complexity" of a network (especially in the context of Activity Networks, which we presume are his prime interest). However, it seems fair to conclude from his article that he equates complexity to reduced parallelism among the paths of the network, and conversely, that simplicity is therefore equalled to parallelism. (see last paragraph on page 172). The rationale behind this conclusion is not evident to us. Parallelism of paths may, or may not, be a blessing depending on the objective of analysis. Furthermore, parallelism itself is not a well-defined notion, and it is of little help to refer to it as a point of reference.

(2) The author insists that the measure of complexity he proposes is useful to "first time users". Here we are at a loss at understanding the author's objective and the intent of his statements. Does the "complexity" of a network vary with the experience of the analyst, so that a network may be both "simple" and "complex" depending on the person doing the analysis ? Or is it a property of the network relative to other networks, which is invariant with the analyst, though the scale on which it is measured may differ ?

(3) In spite of the author's avowed respect for parallelism, or the lack of it, as the determining factor of the network's complexity, his "coefficient of network complexity, (CNC)" is devoid of serious con-

sideration of that factor. For he defines CNC = A^2/N; where A is the number of arcs and N the number of nodes in the network. This CNC, together with all similar measures proposed by others (see below), almost ignore the concept of parallelism. To see this, consider the two simple networks shown in Figure 1. Both networks have A=5 and N=4. Thus, both yield the same CNC ! And yet, if the objective is to determine the probability distribution function of node 4, the network in (b) is much more difficult to analyze than that of (a).

(4) The author offers three possible uses of his (or anybody else's) CNC. They may be summarized as follows (see "Potential Value", p.176-177) :

(i) As a measure of the "... degree of attention that has been devoted to the planning of the project";

(ii) As a predictor of "... the processing time requirements for a particular software package for a particular piece of hardware";

(iii) As an evaluator of proposed algorithms for the analysis of the network.

Once more we feel at a loss in understanding, or accepting, uses (i) and (iii). We submit that it is impossible to look at a network and derive a number (or a set of numbers) which indicate the "degree of attention" paid to the construction of the network. The "simplicity" or "complexity" of a network - in whichever terms these words are defined - are dependent on the project itself, and not on the "attention" paid by the analyst. In other words, a "simple" network is a reflection of a "simple" project, and not of a poor analyst !

As to objective (iii) it is difficult for us to see how a "measure of complexity", by itself, unassisted by other information, can lead to the rejection of an algorithm as inappropriate for the analysis of a network.

Finally, objective (ii) seems to us to be a valid one. But then it is a relative measure of an intrinsic property of the project, which contradicts the assumption of its variability with the analyst (see (2) above) !

(5) Regrettably the author has ignored two measures of network complexity that have been proposed by other authors, and which seem to be in the same spirit as his ; viz., the measures of Davies [1] and Pascoe [4] (This latter measure was adopted by Davis [2]). They are :

$$\text{CNC-D} = 2(A-N+1)/(N-1)(N-2) \qquad \text{(Davies)}$$
$$\text{CNC-P} = A/N \qquad \text{(Pascoe)}$$

In our view, these measures suffer from the same logical flaws that plague the author's measure (see General Comments below), albeit they possess the same "elegance of simplicity" ([3] , p. 173) claimed by the author for his CNC.

(6) Regrettably, the author's figures shown in his Table 1 are in error. In particular, the number of activities listed there do not coincide with the actual count of activities from the networks themselves. (1(a) has 35 activities, not 23 ; 1(c) has 31 activities, not 34 ; and 1(d) has 36 activities, not 40). This rendered the correspon-ding values of CNC-K of his Table 1 wrong.

(7) For the sake of comparison, and to illuminate our subsequent discus-sion, we decided to determine the CNCs of the two other authors for the six networks given by Kaimann (his Figures 1(a) - 1(f)) - with the corrected A and N. These are given in Table 1 below. The actual CPU-time, as reported by Kaimann, is shown in column 3. For each author, Table 1 presents three columns. The first is the CNC according to the author's formula. The second is the standardized-CNC for which the origin (value = 0) is at the lowest CPU-time (occurred at network 2 for all three measures), and the value 10 is

at the highest CPU-time (occurred at network 6 for all three measures). The "standardization factor" is the difference between the highest and the lowest CNCs; also shown in Table 1 for each coefficient. Finally, the third column gives the value of the least squares CPU-time estimate of the quadratic equation passing through the given six points : $y = a_o + a_1x + a_2x^2$. The standardized CNCs are plotted vs. the CPU-time in Figure 2.

Two interesting conclusions emerge. First that everybody's coefficient yields a quadratic function when applied to these networks. Notice also the very high coefficients of determination (corrected for the degrees of freedom) shown in Table 1. This is to be expected with so few points. Second, that Kaimann's coefficient yields slightly higher estimates than the other two (which yield almost identical results in spite of their radically different expressions). But since interest lies solely in the qualitative characterization of the relationship and not in its precise functional form, we conclude that all three measures assert the same result.

We are thus forced to ask the question : to what end does a "new" CNC serve ?

We propose now to take a more general view of all these "coefficients", in the hope of setting the stage for more fundamental contributions to this issue.

II. GENERAL REMARKS.

It has been previously remarked by other researchers that all measurement starts with a sensation. The physical sciences are replete with examples illustrating this point : one feels cold or hot and one inquires about the measurement of heat ; or one feels that a substance is more amenable to elongation than another and inquires about its elasticity ; or one feels that a fluid is more "sticky" than another and inquires about its viscosity ;

and so forth. Indeed, modern technology attests to human triumph in according measure to these sensations and, more importantly, in manipulating the factors that determine their magnitudes.
Therefore there is no doubt in our minds about the importance of the problem to which the author, and others, have addressed themselves. Questions of measurement lie at the very foundation of all scientific progress.

Sometimes the desired measure turns out to be as simple as a multiple of an arbitrary number - such as the measure of length. Sometimes it is a vector of two numbers, such as the measure of heat which requires quantity and potential (i.e. temperature), and which is not additive in any direct sense but involves the even more subtle concept of entropy to effect such "addition". Moreover, the measure can be a complicated expression combining several factors and "universal" constants, such as Young's modulus of elasticity. Oftentimes the measure is unabashadely arbitrary, albeit standard, such as the measure of surface hardness (which is measured by either the diameter of indentation made by a hardened steel sphere -- the Brinnell scale -- or the height of rebound of a small drop hammer -- the Shore Scleroscope scale).

Apart from such well-known measures of physical entities, there have been recently some giant steps in the measurement of non-physical sensations such as intelligence, aptitude, perception, etc. The most outstanding examples that come to mind in this regard are two : the measurement of utility and the measurement of the information content of a message. In each case a sensation was translated into a numerical value. The measures derived are not intuitively apparent, or immediately derivable, from the original sensation, being heavily dependent on concepts of probability which, a priori, did not seem to play any role.

We do not wish to transform this note into an essay on the theory of measurement. The interested reader may wish to consult ref. [5]. But we would be remiss not to highlight the difference between the approaches to the construction of measures, viz., the deductive approach versus the inductive approach. On the one hand, the inductive approach depends on vast amounts of observation and/or experimentation, first to suggest the measure and then to verify its validity (i.e., that it measures what it is supposed to measure) and its reliability (i.e., the consistency of the measures obtained by different observers). On the other hand, the deductive approach relies on the characterization of a set of axioms (or desiderata) which the desired measure should satisfy. This leads to either the choice of a measure from among several, or to the determination that the axioms are satisfied by one and only one measure,or to the assertion that no measure exists which satisfies the stated axioms.

Returning now to the various papers which propose measures of network complexity, we are vexed by the nonchalant manner in which these measures have been offered. There has been neither observation of the phenomenon nor axiomatization of the properties of the measure. Worse still, there has not been an attempt to enumerate the factors that affect the sensation (except, of course, for the obvious values of A and N). For all the measures offered thus far the issue of reliability has been guaranteed to anyone who can count nodes and arcs ; but there has been no attempt at validation. The literature is conspicuously void of either analytical proofs or of any well-designed experiment that verifies any property of the proposed measures - not even the question of whether they are ordinal or cardinal measures ! (In other words, is a network with CNC=4 twice as complex as a network with CNC=2 ?)

All the measures we are aware of lack the basic requisite of a measure, namely, that a different sensation corresponds to a different measure. This is evident from their application to the two networks of Figure 1.

We submit that the "complexity" of a network is a sensation that is inexorably entwined with the objective of analysis. Thus a network may be simultaneously "complex" relative to one objective and"simple"relative to another. Hence such objective must constitute the starting point of defining the sensation and of any subsequent attempt at its measurement.

To illustrate, we propose two measures for two different objectives of analysis.†

Objective 1 : To calculate the length of the Critical Path. In this instance it seems logical to use the "number of comparisons" as the independent variable, since the effort expended in such determination is directly proportional to the number of comparisons. It is easy to deduce that this latter is equal to A-N+1 ; i.e. for this objective,

$$(CNC)_1 = A-N+1 \ .$$

The reader may verify that the two networks of Fig. 1 do indeed utilize exactly the same number of comparisons, namely 2 (= 5-4+1). Hence, relative to this objective, the two networks are equally complex.

Objective 2 : To Determine the Probability Distribution Function of the Terminal Event. Here we assume that the activity durations are random variables with known p.d.f.'s. For this objective we propose the following entity as the independent variable : The minimum number of arcs to be "conditioned on" in order to reduce the network to a set of parallel paths. The motivation for this measure is as follows. If all the paths were in parallel, the p.d.f. of the terminal event would be simply the product of the paths' p.d.f.'s. For paths not in parallel, one fixes the durations of some arcs in order to reduce the paths to parallel ones. Multiplying the p.d.f.'s now yields the p.d.f. of the terminal event conditioned on the values of the specified arcs.

One then removes the conditioning through multiple integration over the p.d.f.'s of these arcs. It is this process of "integrating out"

---

† The authors are grateful to Mr. A.H.J. Rinnooy Kan and Mr J.K. Lenstra, both of the Math. Centrum, Amsterdam, Holland, for their valuable stimulation to develop these two measures.

the conditional arcs that "complicates" the determination of the p.d.f. of the terminal node. Hence it is natural to consider the number of the arcs "conditioned on" as the independent variable. We insist on the minimum such number in order to ensure the minimum amount of effort. In other words, we propose the following measure for this objective :

$$(CNC)_2 = g(\nu) ;$$

where ν is the number of the arcs "conditioned on", and g is some function whose exact form may be determined by measuring, say, the CPU-time for different values of ν empirically. In all probability it will be a nonlinear function.

If we apply this measure to the two networks of Fig. 1, we find that $\nu_a = 0$ (since all three paths are in parallel) while $\nu_b = 2$; (i.e., we must "condition" on two arcs in the network (b) in order to obtain a set of three independent paths). We terefore conclude that, <u>for this objective, network</u> 1(b) <u>is more complex than network</u> 1(a) !! The exact quantification of this ordinal statement is available once the form of the function g is determined. Thus, if g is an exponential function of the form ab^{ν} , $b > 1$, then the network in 1(b) would be b^2 times as complex as that in 1(a). Such quantification is currently under investigation.

Comparing this statement with the conclusion under Objective 1, we discover that the network in 1(b) is simultaneously "easy" and "complex", depending on the objective of analysis, as we previously asserted.

Clearly, these two objectives do not exhaust the list of objectives, even within the limited context of Activity Networks - for example; we are not aware of any measure corresponding to the objective of optimal scheduling of activities subject to scarce resources. We therefore conclude that more research is needed to determine other measures that would stand the tests of logic and experimentation.

References.

1. Davies, E.M., "An Experimental Investigation of Resource Allocation in Multiactivity Projects", Operl. Res. Quart. Vol. 24, 4 (1974), 587-591.

2. Davis, E.W., "How Do you Describe a Project Network ?", Paper presented at the ORSA Spring Meeting, 1972.

3. Kaimann, R.A., "Coefficient of Network Complexity", Management Science, Vol. 21, 2 (1974) 172-177.

4. Pascoe, T.L, "Allocation of Resources CPM", Revue Française de Recherche Operationnelle, No. 38, (1966), 31-38.

5. Pfanzagl, J., "Theory of Measurement", J. Wiley & Sons, 1968.

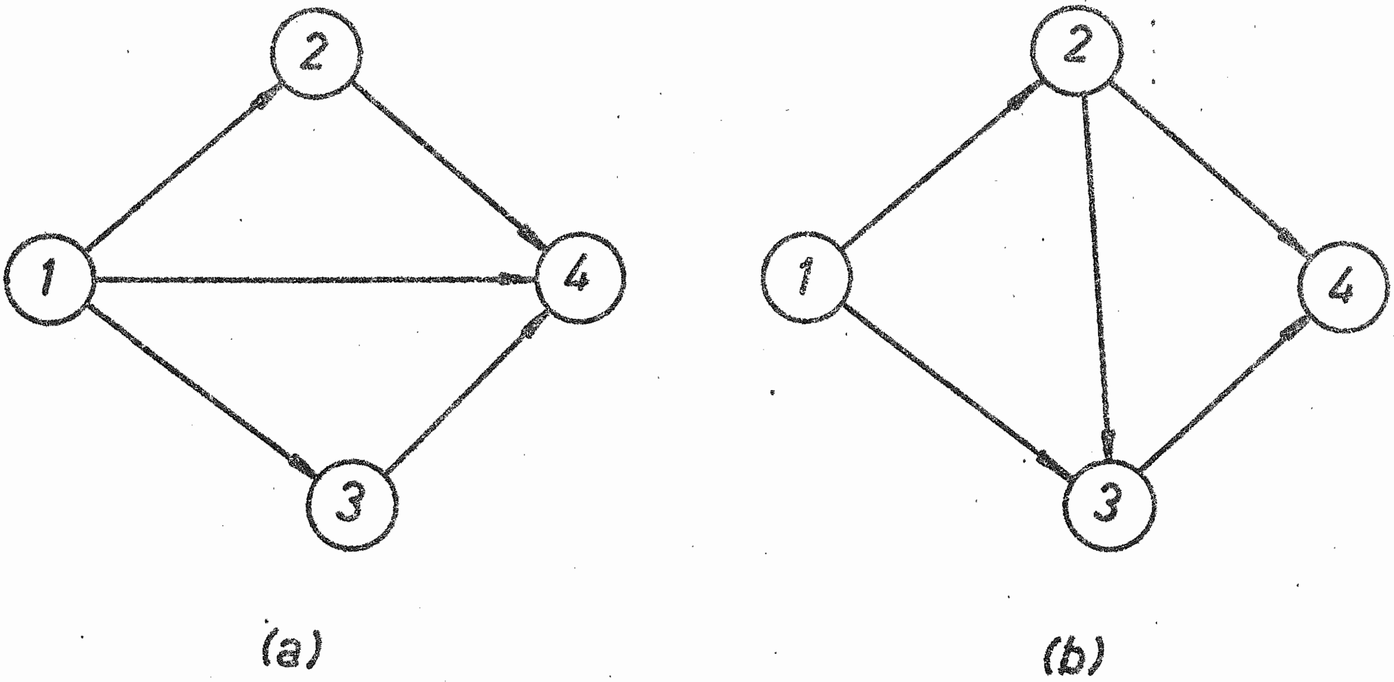
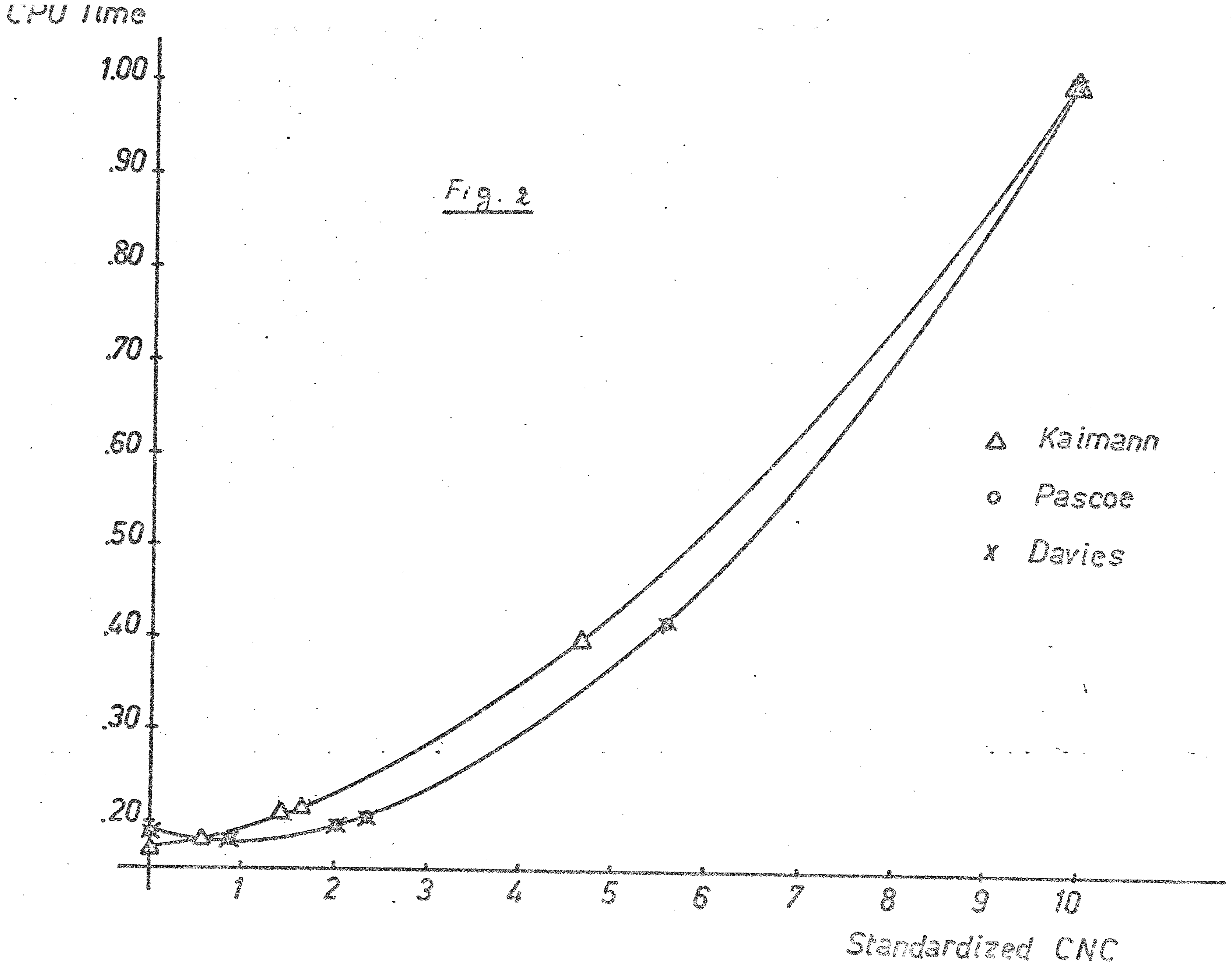


Fig. 1



CPU Time
1.00
.90
.80
.70
.60
.50
.40
.30
.20
1
2
3
4
5
6
7
8
9
10
Standardized CNC
Δ Kaimann
o Pascoe
x Davies

Fig. 2

			KAIMANN			PASCOE			DAVIES		
Network	No. of Activities A	Actual CPU Time Secs	CNC-K	Std. CNC-K	Regressn. y_k	CNC-P	Std. SNC-P	Regressn. y_P	CNC-D	Std. CNC-D	Regressn. y_D
1	35	.15742	55.68	1.44	.2103	1.5909	2.06	.1990	.0067	2.0630	.1991
2	28	.17268	35.64	0	.1732	1.2727	0	.1861	.0333	0	.1861
3	31	.19102	43.68	.58	.1845	1.4091	.88	.1813	.0476	.8833	.1813
4	36	.26832	58.91	1.67	.2190	1.6364	2.35	.2077	.0714	2.3533	.2076
5	47	.39680	100.41	4.66	.3997	2.1364	5.59	.4166	.1238	5.5899	.4166
6	62	1.04017	174.73	10.00	1.0340	2.8182	10.00	1.0357	.1952	10.00	1.0356
		Standardization Factor	139.09			1.5455			.1619		
				Coefficient of Determination	0.9847			0.9824			0.9824

N = 22 for all Networks

CNC-K = $A^2/22$

CNC-P = $A/22$

CNC-D = $(A-21)/210$

Table 1.